ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

VIROLOGY

Virology 311 (2003) 1–6

www.elsevier.com/locate/yviro

Complete nucleotide sequence of an endogenous retrovirus from the amphibian, *Xenopus laevis*

Roziah Kambol,[1] Peter Kabat,[2] and Michael Tristem*

Department of Biological Sciences, Imperial College, Silwood Park, Buckhurst Road, Ascot, Berkshire, SL5 7PY, UK

Received 2 December 2002; returned to author for revision 14 January 2003; accepted 3 March 2003

Abstract

We report the first full-length sequence of an endogenous amphibian retrovirus derived from the African clawed toad *Xenopus laevis*. The virus, termed Xen1, has one of the largest endogenous retroviral genomes described to date of over 10 kb in length and it also has a relatively complex genomic organisation consisting of LTR-*orf1*, *orf2*, *gag*, *pol*, *env*-LTR. Orfs 1 and 2 are novel, duplicated genes of unknown function. Phylogenetic analysis indicates that Xen1 is most closely related to the ε-retroviruses WDSV and WEHV types 1 and 2, which are large, complex exogenous retroviruses present within Walleye fish.

Keywords: Retrovirus; African clawed toad; Genus; Amphibian; Endogenous; Duplication

Introduction

Members of the *Retroviridae* are present within the genomes of all but the most basal vertebrates (Herniou et al., 1998). They are currently classified into seven genera, six of which comprised viruses largely confined to mammals and birds, whereas members of the seventh (the ε-retroviruses) are present within piscine hosts (van Regenmortel et al., 2000). There are currently few full-length retroviral sequences with hosts from other vertebrates, although two recent reports have described the first full-length endogenous retroviruses harboured by several species of reptiles (Huder et al., 2002; Martin et al., 2002). Many of these viruses have unique features of genomic organisation when compared to the retroviruses present within mammals and birds (van Regenmortel et al., 2000).

Here we report the characterisation of an endogenous amphibian retrovirus termed Xen1. This element has an unusual genomic organisation and phylogenetic analyses indicate that it clusters with the ε-retroviruses Walleye dermal sarcoma virus (WDSV) and Walleye epidermal hyperplasia virus (WEHV) types 1 and 2.

Results and discussion

The Xen1 provirus was found to be 10,207 bp in length, making it one of the largest proviral genomes described. Only the spumaretroviruses, Snakehead retrovirus (SnRV at 11.2 kb), HERV.FRD (human endogenous retrovirus type FRD at 10.8 kb), and the three retroviruses from Walleye fish (all approximately 13 kb) are longer (Hart et al., 1996; Holzschu et al., 1995; LaPierre et al., 1999; Petropoulis, 1997; Tristem, 2000). It has a genomic organisation consisting of LTR-*orf1*, *orf2*, *gag*, *pol*, *env*-LTR (see Fig. 1) and is therefore relatively unusual for an endogenous retrovirus, in that it has two extra genes in addition to *gag*, *pol*, and *env*. Other endogenous retroviruses with additional gene products are HERV.FRD, which has a putative 2 kb gene located 5′ of *gag* (Tristem, 2000), and several dUTPase-containing viruses such as HERV.L, HERV.K10, and mouse mammary tumour virus (MMTV). In these cases the dUTPase is located either 3′ of the *integrase* or *protease*

* Corresponding author. Fax: +020-75942339.
E-mail address: m.tristem@imperial.ac.uk (M. Tristem).
[1] Present address: School of Science and Technology, Universiti Malaysia Sabah, Locked Bag No 2073, Kota Kinabalu 88999 Sabah, Malaysia.
[2] Present address: Institute of Virology, Slovak Academy of Sciences, Dubravska cesta 9, 84246 Bratislava, Slovak Republic.

0042-6822/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0042-6822(03)00263-0

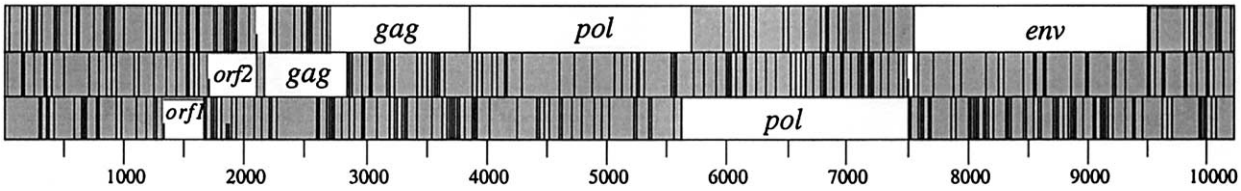

Fig. 1. Open reading frame map of Xen1. Full-length bars indicate stop codons; half-length bars indicate initiation codons. Proviral gene regions are unshaded and labelled. There are four frameshift mutations in the sequence; two are approximately in the centre of the *gag* and *pol* genes, whereas the other two are at the 5′ ends of *gag* and *env*.

genes within *pol* (Cordonnier et al., 1995; Petropoulis, 1997). MMTV also has a superantigen (*sag*) gene situated 3′ of *env* (Petropoulis, 1997). Xen1, although an endogenous virus, is almost intact, having only four frameshift mutations and no obvious in-frame stop codons (see Fig. 1).

To estimate the approximate integration date of Xen1-like viruses, we performed Southern hybridisation analysis on several members of the *Xenopus* genus. This showed that Xen1-like viruses are widespread in the *Xenopus* genus, being located in species such as *Xenopus laevis poweri*, *X.l. sudanensis*, *X.l. victorianus*, *X. fraseri*, *X. muelleri-east*, *X. borealis*, and *X. tropicalis* (R. Kambol, unpublished data). The origin of extant members of the *Xenopus* genus has been estimated at 48 mya, suggesting that Xen1-like viruses have been active in the *Xenopus* lineage for much of this time (Kobel et al., 1998).

The 5′ and 3′ LTRs of Xen1 are 550 and 548 bp in length, respectively, and contain recognisable promoter and polyadenylation signals (see Fig. 2). The primer binding site (PBS) of Xen1 is complementary to $tRNA^{Lys2}$ (17 of 18 bp matches) and is identical to the PBS of Mason–Pfizer monkey virus (MPMV) (Sonigo et al., 1986). The LTRs are more than 99% identical to each other, which is consistent with the almost intact nature of the virus, and suggest that Xen1 itself has only recently integrated into the genome of its host.

Xen1 has a large leader region of over 750 bp, which contains several repeats of approximately 40 bp in length, and this region is followed by two relatively short ORFs encoding protein products of 109 and 132 amino acids, respectively. The two ORFs appear to have arisen via gene duplication, as they are 36% identical at the amino acid level (Fig. 3). A variety of databank searches failed to reveal any similarity of these ORFs to other proteins (Altschul et al., 1997), and they do not appear to contain motifs suggesting any particular functions (M. Tristem, unpublished data). Three previous examples of gene duplication within retroviruses have been reported. The *vpx* gene in the lentiviral HIV-2 group appears to have arisen via duplication of *vpr*, whereas the lentiviral *rev* gene may have arisen via partial *env* gene duplication (Kubota et al., 1994; Tristem et al., 1992). In the former case, the original (unduplicated) *vpr* gene had several functions, and these became split in the descendent, duplicated genes (Fletcher et al., 1996). The other known duplication occurred within the piscine retroviruses WDSV and WEHV1 and 2, in which the duplicated genes encode cyclin D homologs (LaPierre et al., 1998, 1999). This last example is the only case where a retroviral accessory gene has a known cellular origin.

Xen1 *gag* is 1.8 kb in length and contains a consensus myristylation signal at its 5′ end, a major homology region (MHR) within CA and, similar to certain retroviruses such as human immunodeficiency virus type-1 (HIV-1) and Avian sarcoma/leukosis virus (ASLV), encodes two Cys-His boxes within nucleocapsid (NC) (Vogt, 1997). Synthesis of the Gag-Pol polyprotein appears to be mediated via termination suppression, as is the case with MLV (Swanstrom and Wills, 1997). Xen1 encodes an amber termination codon at the 3′ end of *gag*, which is followed immediately by a sequence predicted to form a hairpin loop and associated pseudoknot structure (R. Kambol, unpublished data). Pol extends for 3.8 kb and has the typical retroviral organisation of Pro, RT/RNaseH, Int, and no other gene products.

Env was found to be approximately 2 kb in length and BLAST searches revealed similarity to several γ-retroviruses, such as Gibbon ape leukaemia virus (GaLV) and Baboon endogenous retrovirus (BaEV). There was no obvious sequence similarity with any lower vertebrate derived retroviruses, including WDSV and WEHV types 1 and 2 (M. Tristem, unpublished data).

A phylogeny, based on a 350 amino acid alignment of conserved regions of retroviral reverse transcriptase (RT), RNase H, and integrase (Int) proteins, was next constructed (as in Martin et al., 2002). This data set included all of the full-length reptilian and piscine retroviral sequences available in GenBank, as well as representative examples of mammalian and avian retroviruses. It was apparent that although Xen1 was distinct from other previously described retroviruses, there was some bootstrap support for a clade containing Xen1, WDSV, and WEHV1 and 2 (see Fig. 4). The other member of the ε-retroviruses, SnRV, was not placed within this clade.

Several features of the genomic organisation of the Walleye viruses and Xen1 are consistent with their placement as sister taxa. Both Xen1 and the Walleye derived viruses are large, complex retroviruses encoding several accessory genes in addition to *gag*, *pol*, and *env*. All these viruses also encode Gag and Pol in the same ORF frame and hence use termination suppression rather than ribosomal frameshifting to produce Gag/Pol polyproteins.

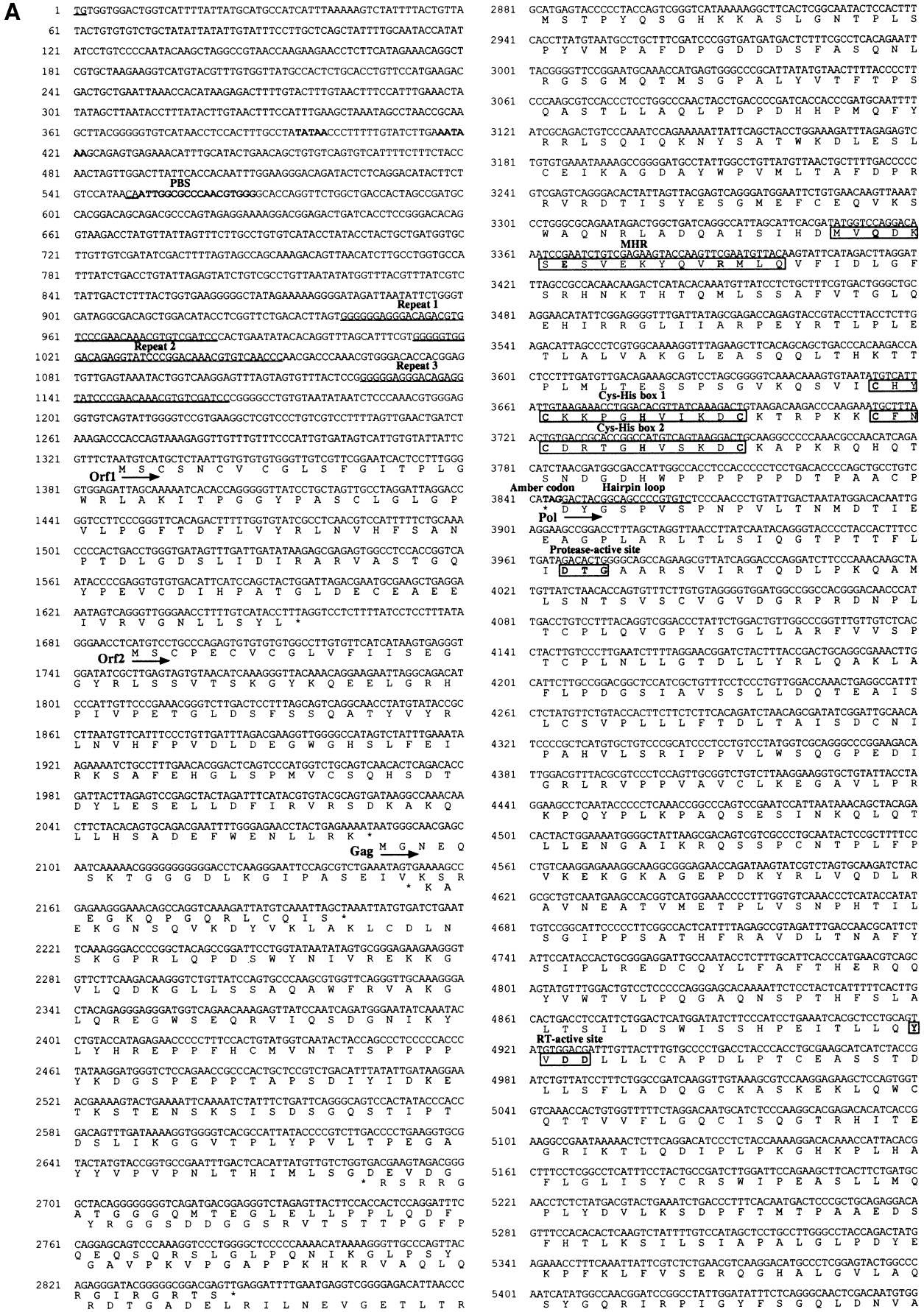

Fig. 2. Complete nucleotide sequence of the Xen1 provirus. The inverted repeats at each end of the LTR are underlined, as is the potential hairpin loop at the start of *pol*, and the three 40-bp repeats in the leader sequence. Bold residues indicate the promoter and polyadenylation signals within the 5′ LTR, the PBS and PPT, and the amber termination codon at the end of *gag*. Arrows indicate the 5′ ends of the encoded proteins. Other features of the sequence are boxed and labelled, with highly conserved amino acid residues within each motif further highlighted in bold type.

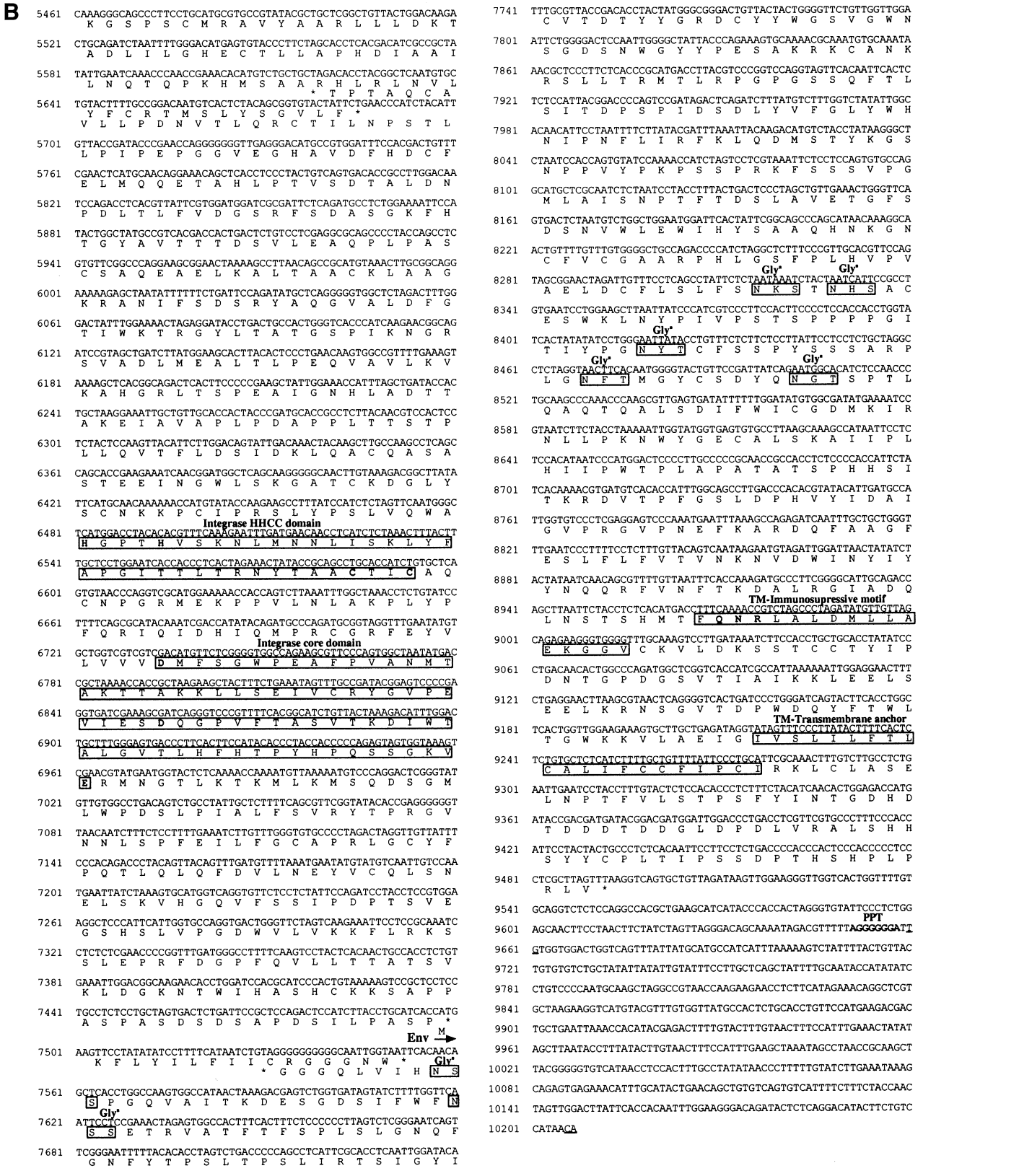

Fig. 2 (*continued*)

However, despite these similarities, Xen1 is still clearly distinct from the piscine retroviruses. For example, the Walleye fish derived viruses contain two accessory genes 3′ of *env* not observed in Xen1, and we were unable to find any obvious homology between any of the accessory genes of Xen1 and the Walleye fish derived viruses. The *env* genes of

```
orf1   1 MSCSNCVCGLSFGITPLGWRLAKITPGGYPASCLGLGPVLP--GFTDF-----LVYRLNV 60
orf2     MSCPECVCGLVFIISEGGYRLSSVTSKGYKQEELGRHPIVPETGLDSFSSQATYVYRLNV

orf1  61 HFSANPTDL----GDSLIDIRARVASTGQYPEVCDIHPATGLDECEAEEIVRV------- 120
orf2     HF---PVDLDEGWGHSLFEIRKSAFEHGLSPMVCSQHSDTDYLESELLDFIRVRSDKAKQ

orf1 121 ---------GNLLSYL 136
orf2     LLHSADEFWENLL-RK
```

Fig. 3. Alignment of the duplicated orf1 and orf2 genes within Xen1. Identical amino acid residues between the two proteins are highlighted in bold. BLAST-to-BLAST searches gave an expect value of 4E−13, indicating that the similarity is not due to chance.

the WDSV and WEHV types 1 and 2 are unusual in that they contain exceptionally long cytoplasmic domains (of at least 200 amino acids), 3′ of the transmembrane anchor in TM (LaPierre et al., 1999). It has been suggested that this may help to stabilise the virus for transmission in aquatic environments (LaPierre et al., 1999). In contrast, Xen1 contains a TM with relatively high sequence similarity to the γ-retroviruses, and hence, has a cytoplasmic tail of approximately 70 amino acids. This is despite being harboured by a host species belonging to the Anuran family *Pipidae* (Clawed and Surinam toads), the members of which are almost entirely aquatic.

Materials and methods

Cloning and sequencing

Xen1 was isolated from a *X. laevis* (African clawed toad) library using genomic DNA obtained from the cell-line HD-1 (Reinschmidt et al., 1985). Sau3A-digested fragments were ligated into the vector-λ EMBL4 and the resultant plaques were probed with the ^{32}P-labelled RV-African clawed toad retroviral fragment described in Herniou et al. (1998). The full-length sequence (both strands) of one clone, termed Xen1, was obtained using the "primer walking" method and an automated sequencer.

The sequence of Xen1 has been submitted to the EMBL/GenBank/DDJ databases with the Accession No. AJ506107.

Phylogenetic analysis

To maximise the amount of phylogenetic signal in the dataset, an alignment was constructed using well-conserved regions of several retroviral proteins; namely RT (220 amino acid residues), Rnase H (41 residues), and Int (89 residues). The alignment is available from the authors on request. Trees were constructed using maximum parsimony with the program PAUP4 and the Protpars matrix. Bootstrap analysis was from 100 replicates using the Protpars matrix. Neighbour-joining bootstrapping was also performed, but in this case the data matrix was subjected to 1000 replicates.

Southern hybridisation

Genomic DNA from seven members of the *Xenopus* genus (*X.l. poweri*, *X.l. sudanensis*, *X.l. victorianus*, *X. fraseri*, *X. muelleri-east*, *X. borealis*, and *X. tropicalis*) was probed with a ^{32}P-labelled *pol* gene fragment of Xen1 using standard conditions (Sambrook et al., 1989). Filters were washed down to 1× SSC/0.5%SDS at 55°C.

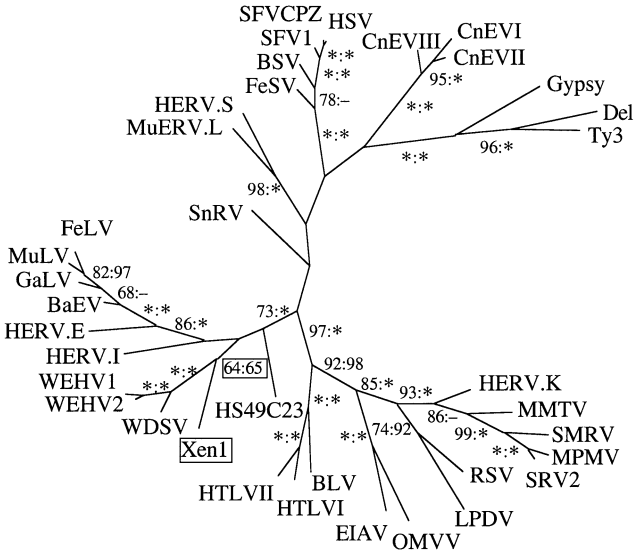

Fig. 4. Unrooted maximum parsimony phylogeny of Xen1 and other members of the *Retroviridae*. Three *Gypsy*-type LTR-retrotransposons were also included as outgroup taxa. Branch lengths are proportional to the level of sequence divergence. Figures indicate bootstrap percentages from 100 maximum parsimony replicates on the left and 1000 neighbour-joining replicates on the right. Asterisks indicate bootstrap percentages of 100.

Acknowledgments

We thank T. Sargent for kindly providing the *Xenopus* genomic library and H. R. Kobel and C. Thiebaud for the DNA samples of the *Xenopus* subspecies. Thanks also to A. Katzourakis, R. Gifford, C. Lynch, and J. Martin for comments. R.K. is indebted to the Malaysian Government for a research studentship.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.L., 1997. Gapped BLAST and PSI-BLAST: a new

generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Cordonnier, A., Casella, J.P., Heidmann, T., 1995. Isolation of novel human endogenous retroviral-like elements with foamy virus-related pol sequence. J. Virol. 69, 5890–5897.

Fletcher, T.M., Brichacek, B., Sharova, N., Newman, M.A., Stivahtis, G., Sharp, P.M., Emerman, M., Hahn, B.H., Stevenson, M, 1996. Nuclear import and cell cycle arrest functions of the HIV-1 Vpr protein are encoded by two separate genes in HIV-2/SIV(SM). EMBO. J. 15 (22), 6155–6165.

Hart, D., Frerichs, N., Rambaut, A., Onions, D.E., 1996. Complete nucleotide sequence and transcriptional analysis of the snakehead fish retrovirus. J. Virol. 70, 3606–3616.

Herniou, E., Martin, J., Miller, K., Cook, J., Wilkinson, M., Tristem, M., 1998. Retroviral diversity and distribution in vertebrates. J. Virol. 72, 5955–5966.

Holzschu, D.L., Martineau, D., Fodor, S.K., Vogt, V.M., Bowser, P.R., Casey, J.W., 1995. Nucleotide sequence and protein analysis of a complex piscine retrovirus, walleye dermal sarcoma virus. J. Virol. 69, 5320–5331.

Huder, J.B., Böni, J., Hatt, J.-P., Soldati, G., Lutz, H., Schüpbach, J., 2002. Identification and characterisation of two closely related unclassifiable endogenous retroviruses in pythons (*Python molurus* and *Python curtis*). J. Virol. 76, 7607–7615.

Kobel, H.R., Barandun, B., Thiébaud, C.H., 1998. Mitochondrial rDNA phylogeny in *Xenopus*. Herpetological J. 8, 13–17.

Kubota, S., Oroszlan, S., Hatanaka, M., 1994. The origin of human immunodeficiency virus type-1 *rev* gene. An evolutionary hypothesis. FEBS Lett. 338 (2), 118–121.

LaPierre, L.A., Casey, J.W., Holzschu, D.L., 1998. Walleye retroviruses associated with skin tumors and hyperplasias encode cyclin D homologs. J. Virol. 72 (11), 8765–8771.

LaPierre, L.A., Holzschu, D.L., Bowser, P.R., Casey, J.W., 1999. Sequence and transcriptional analyses of the fish retroviruses walleye epidermal hyperplasia virus types 1 and 2: evidence for a gene duplication. J. Virol. 73 (11), 9393–9403.

Martin, J., Kabat, P., Herniou, E., Tristem, M., 2002. Characterization and complete nucleotide sequence of an unusual reptilian retrovirus recovered from the order Crocodylia. J. Virol. 76, 4651–4654.

Petropoulis, C., 1997. Retroviral taxonomy, protein structures, sequences, and genetic maps, in: Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.), Retroviruses, CSHL Press, New York, pp. 757–805.

Reinschmidt, D., Friedman, J., Hauth, J., Ratner, E., Cohen, M., Miller, M., Krotoski, D., Tompkins, R., 1985. Gene-centromere mapping in *Xenopus laevis*. J. Hered. 76 (5), 345–347.

Sambrook, J., Fritsch, E., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sonigo, P., Barker, C., Hunter, E., Wain-Hobson, S., 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell 45, 375–385.

Swanstrom, R., Wills, J.W., 1997. Synthesis, assembly, and processing of viral proteins, in: Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.), Retroviruses, CSHL Press, New York, pp. 263–334.

Tristem, M., 2000. Identification and characterisation of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. J. Virol. 74, 3715–3730.

Tristem, M., Marshall, C., Karpas, A., Hill, F., 1992. Evolution of the primate lentiviruses: evidence from *vpx* and *vpr*. EMBO J. 11, 3405–3412.

van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., Wickner, R.B., 2000. Virus taxonomy: the classification and nomenclature of viruses. The seventh report of the International Committee on taxonomy of viruses. Academic Press, San Diego.

Vogt, V.M., 1997. Retroviral virions and genomes, in: Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.), Retroviruses, CSHL Press, New York, pp. 27–69.